MULTIVARIATE METHODS FOR SIZE-DEPENDENT DETECTION IN CONVENTIONAL LINE TRANSECT SAMPLING

Jaume Forcada[1]

Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA
8604 La Jolla Shores Drive
La Jolla, California 92037

JUNE 2002

ADMINISTRATIVE REPORT LJ-02-07

---

[1] Present address:
NERC, Biological Sciences Division, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
e-mail: j.forcada@bas.ac.uk

ABSTRACT

When animals occur in groups, biases in mean group size estimates bias abundance estimates using conventional line transect methods. Multivariate models for bias correction of mean group size and for detection function analysis in conventional line transect are reviewed, developed and tested. Standard bias-correction methods, based on a least squares analysis of the observed log-transformed group sizes against estimated detection probability, tend to underestimate mean group size when outliers are present. In contrast, robust regression improves mean group size and abundance estimates when the association of perpendicular distances with other covariates is linear. Parametric and nonparametric multivariate detection function models incorporated into line transect abundance estimators provide substantial improvement in estimating mean group size and abundance. Under the assumption of perfect detection in the transect line, $g(0)=1$, the flexibility of multivariate models to fit the detection function near 0 is critical to obtain unbiased abundance estimates. Nonparametric methods usually have a poor fit of the short perpendicular distances and produce less precise estimates as the number of covariates increases. Thus, there is a need for better smoothing algorithms and also for an objective covariate selection to improve abundance estimates. Parametric methods, allowing for objective model selection and averaging, provide the best trade-off of bias and precision, and produce more reliable estimates of abundance and its variability when sample size is not low.

INTRODUCTION

Conventional line transect sampling is commonly used to estimate abundance, a quantity of fundamental interest in ecological studies. Given a study area and survey design, the main data are right angle distances to objects detected along a transect line. The term conventional applies when detection of objects directly on the line is assumed perfect, and many studies rely on this assumption. When detected objects are groups, the analysis proceeds by estimating the effective sampling distance on each side of the line, and density is the product of the inverse of this estimate, the groups encounter rate and the mean group size (Buckland et al., 2001). This analysis, however, can be problematic when group size is highly variable.

Sampling wildlife populations, large groups provide more visible cues and are detected at larger perpendicular distances than small groups. This causes a selection bias in the observed group size distribution, and the estimated mean is unrepresentative of the true mean group size in the survey area. Moreover, the detection of these groups is usually a heterogeneous process, significantly affected by sighting conditions, observers training and behavior, the changing environment, and occurrence in mixed-species groups. For instance, a small single-species group within a larger mixed-species group is easier to detect than it otherwise might be. Typical examples are multispecies aggregations of African ungulates, schooling fish, or oceanic dolphins. Under these conditions, heterogeneity may not be accounted for by conventional methods, based on perpendicular distance models only. In this paper, I consider methods for size-dependent detection accounting for multiple covariates in the estimation of density, abundance and expected group size. I illustrate the methods with an example and simulate data to test and compare the performance of the proposed and other available methods.

The problem of size-dependent detection has been addressed with linear regression methods (Buckland et al., 2001). Modeling the $\log_e$ of the observed group size as a function of detection probabilities, the expected group size can be predicted at the transect line where detection is assumed perfect. I extend this method to account for multiple covariates affecting the group size distribution and I develop a more general estimator to account for mixed-species groups. Further, I cast the estimator into a robust-regression context, resistant to outlier effects and non-Gaussian error distribution.

Size-dependent detection has also been addressed with bivariate detection functions (e.g., Drummer and McDonald, 1987; Quang, 1991; Chen, 1996). The detection probability is estimated by modeling the conditional distribution function of perpendicular distances and group sizes. The estimated bivariate density function evaluated at 0 is then multiplied by the expected group size and the observed sighting rates to estimate density. The generalization of this method, incorporating multiple covariates in the detection function, reduces heterogeneity and deals with size-dependent detection simultaneously. The new estimation framework allows for parametric and nonparametric detection functions, and I propose and test suitable estimators for each case.

I investigate the new methods with line transect data of the eastern spinner (*Stenella longirostris orientalis*) dolphin stock in the eastern tropical Pacific Ocean (ETP). The U.S. National Marine Fisheries Service collected the data during the period 1998-2000 as part of the studies directed by the 1997 International Dolphin Conservation Program Act (IDCPA, 16 U.S. Code 1414), an amendment to the Marine Mammal Protection Act (MMPA, 16 U.S. Code 1361 *et seq*). This law required studies to determine if the chase and encirclement of dolphins in the purse-seine fishery for yellow fin tuna in the ETP is having a significant adverse impact on depleted dolphin stocks, and line transect abundance estimates of dolphin stocks were an integral part of the studies mandated by the U.S. Congress. The affected dolphin stocks are distributed in mixed-species schools ranging in size from a few individuals to thousands of animals, and have a unique association with yellow fin tuna and seabirds (Perrin, 1969). Thus, the detection process is significantly influenced by the size of schools, detection cues and inter-specific associations. In this paper, only a sub sample of the data is analyzed to illustrate the methods presented. The analysis of the complete data set will be presented elsewhere.

Reviewers from the Center for Independent Experts provided an independent peer review of this work. Responses to reviewers' comments can be found in Appendix A.

UNIVARIATE LINE TRANSECT ANALYSIS

In conventional methods, transect lines are defined to have an arbitrary maximum strip width W. Objects observed beyond W are either ignored or discarded for the purpose of density estimation. Given W and L, a transect line of known length, the area surveyed is $a_w = 2LW$ and contains a number of objects N_w. By assumption, N_w is a random variable with expectation $2LWD$, and D, density, is the number of objects per unit area. Any object within a_w is located at a right angle (perpendicular) distance Y from the transect line, and Y is assumed uniform $I(0, W)$. Objects in the strip W are observed with unconditional probability

$$P = \frac{1}{W}\int_0^W g(y)dy = \frac{1}{W}\mu$$

(Seber, 1982; Burnham and Anderson, 1976), where $g(y)$ is the probability that an object is detected given that it is at an observed perpendicular distance y. The expected number of detected objects is

$$E(n) = E(N_W P) = 2LD\mu$$

and density of objects is estimated as

$$\hat{D} = \frac{n}{2L\hat{\mu}}. \tag{1}$$

Detection of objects on the transect line is assumed certain, i.e. $g(0)=1$, and μ, known as the effective strip half-width, is estimated as $\hat{f}(0)^{-1}$, the inverse of the probability density (*pdf*) of observed perpendicular distances y evaluated at 0. When detected objects are groups, density of individuals is estimated as

$$\hat{D}_{ind} = \frac{\hat{E}(s)\, n\, \hat{f}(0)}{2L}, \tag{2}$$

where $\hat{E}(s)$ is the expected mean group size. In practice, $\hat{E}(s)$ is replaced by an estimate of mean group size, assuming that it is independent of the density of observed and unobserved groups. Alternatively, under the same assumptions, a size-bias corrected $\hat{E}(s)$ is estimated by linear modeling of the $\log_e$ of the observed group size and detection probability, $\hat{z} = \log_e(s) = \alpha + \beta\hat{g}(y)$ (Buckland et al., 2001, p. 74-75). Conditional on the regression slope (β) being significant, $\hat{E}(s)$ is predicted at the transect line where $\hat{g}(0) = 1$

$$\hat{E}(s) = e^{\alpha + \beta + \hat{\mathrm{var}}(\hat{z})/2} \tag{3}$$

where

$$v\hat{a}r(\hat{z}) = \hat{\sigma}_\varepsilon^2 \left[1 + n^{-1} + (1-\overline{g})^2 \left\{ \sum_{i=1}^{n} \left(\hat{g}(y_i) - \overline{g} \right)^2 \right\}^{-1} \right], \tag{4}$$

$\hat{\sigma}_\varepsilon^2$ is the residual mean square, and $\overline{g} = n^{-1}\sum_n \hat{g}(y_i)$. Coefficients α and β are estimated with the least-squares (*LS*) method.

Robust-regression estimator of $\hat{E}(s)$

The $\log_e$ of the total group size T, in terms of group size (s) and proportion (r) of the species of interest, $T = s\,r^{-1}$, is regressed against $g(y)$ and a vector of l additional covariates, $z = \log_e(T) = \alpha + \sum_{j=1}^{l} \beta_j x_j$. The expected group size is estimated as

$$\hat{E}(s) = n^{-1} \sum_{i=1}^{n} r_i\, e^{\left\{\alpha + \sum_{j=i}^{l} \beta_j x_{ij} + \frac{\mathrm{v\hat{a}r}(\hat{z})}{2}\right\}} \tag{5}$$

with

$$\mathrm{v\hat{a}r}(\hat{z}) = \hat{\sigma}_{\varepsilon}^{2}\left[1 + n^{-1} + (x^{+})' \left\{(X^{+})' X^{+}\right\}^{-1} (x^{+})\right], \tag{6}$$

where X is the matrix of covariate values, the prime denotes transpose, and X^{+} is the matrix of centered (mean corrected) covariates. If total group size (T) depends on $g(y)$ only, Eq. *5* becomes

$$\hat{E}(s) = n^{-1} \sum_{i=1}^{n} r_i\, e^{\alpha + \beta + \mathrm{v\hat{a}r}(\hat{z})/2} \tag{7}$$

and $\mathrm{v\hat{a}r}(\hat{z})$ is estimated as in Eq. *4*.

The variance of $\hat{E}(s)$ is

$$\mathrm{v\hat{a}r}\left\{\hat{E}(s)\right\} = \left\{R(n-1)\right\}^{-1} \sum_{i=1}^{n} r_i \left[r_i e^{\left\{\alpha + \sum_{j=i}^{l} \beta_j x_{ij} + \mathrm{v\hat{a}r}(\hat{z})/2\right\}} - \hat{E}(s) \right]^2 ,$$

where $R = \sum r_i$, and $\hat{E}(s)$ is assumed independent of density of groups.

Detections of groups located in the same transect are likely to be interdependent and the empirical variance estimate will likely be biased in those cases. More reliable variance and confidence intervals can be estimated with a nonparametric bootstrap resampling from multiple transects, provided that transects are independent (Buckland et al., 2001). Resampling detected groups by transect with replacement allows for an estimate of $\hat{E}(s)$ in each of B bootstrap replicates. $\mathrm{v\hat{a}r}\left\{\hat{E}(s)\right\}$ is then the variance of the B replicates.

Estimation method**.** Coefficients α and β_j in Eqs. *3*, *5* and *7* can be estimated with the *LS* method, assuming that the $\log_e$ transformation of group size makes the data linear with normally distributed errors. However, datasets with highly variable group sizes contain significant outliers with strong influence on the *LS* fit, and robust regression is likely to produce more reliable $\hat{E}(s)$ estimates. I choose the *MM* robust regression method of Yohai et al. (1991) because: *i*) the model fit is minimally influenced by outliers in the response and predictors' space; *ii*) the fit minimizes the bias in coefficients estimates due to non-Gaussian errors; and *iii*) statistical inference is based on large sample size approximations and is comparable to that obtained with the *LS* method.

Given a linear model with the general form $s_i = x_i^T \beta + \varepsilon_i$, a robust *M*-estimate (Yohai et al., 1991) of $\hat{\beta}$ is obtained by minimization of

$$\sum_{i=1}^{n} \rho\left(\frac{s_i - x_i^T \hat{\beta}}{\hat{S}}; c\right),$$

where $\rho(.,c)$ is a convex weight function (e.g. ρ =*-log f*, where *f* is a density function) with $\hat{S}$, a robust scale estimate of the residuals. Because this minimization can have multiple results, an initial estimate $\hat{\beta}^0$ is obtained from a robust regression *S*-estimate (Rousseeuw and Yohai, 1984). $\hat{\beta}^0$ is the value minimizing the robust scale estimate $\hat{S}(\beta)$ in the linear model

$$\frac{1}{n-q} \sum_{i=1}^{n} \rho\left\{\frac{s_i - x_i^T \hat{\beta}}{\hat{S}(\beta)}\right\} = 0.5,$$

and $\hat{\beta}^0$ is used as a local minimum of the $\rho(.,c)$ function of the *M*-estimator to compute the final estimate $\hat{\beta}^1$, and *q* are the smallest squared residuals.

Yohai et al. (1991) developed a test for bias of the final *M*-estimate and a *LS* estimate against the initial *S*-estimate. If the test is significant for the *M* and *LS* estimates, inference with these methods will be biased. Thus, when analyzing a data set, the test for bias of the *M*-estimate should be used and, if not significant, check for a significant regression slope. With a significant slope, use the $\hat{\beta}^1$ estimates to compute $\hat{E}(s)$. If the *M*-estimate is biased but the *LS*-estimate is not, the *LS*-estimate should be used if the corresponding slope is significant. Otherwise, use the observed mean group size.

Covariates explaining significant variation of $\hat{E}(s)$ can be included objectively in the model according to the smallest *RFPE* (Robust Final Prediction Error) of Yohai (*cf* MathSoft Inc., 1999, p. 287)

$$RFPE = \sum_{i=1}^{n} E\left\{\rho\left(\frac{s_i^* - x_i^T \hat{\beta}^1}{\rho}\right)\right\},$$

where s_i^* are the predicted values using the final *M*-estimate, $\hat{\beta}^1$. This statistic can be used as the Akaike's Information Criterion (AIC) (Akaike, 1973), and thus an efficient bootstrap algorithm to estimate the variance of $\hat{E}(s)$ can include covariate selection and test for bias on each replicate.

MULTIVARIATE LINE TRANSECT ANALYSIS

The multivariate detection function $g(y, \underline{c})$ is defined as the conditional probability of sighting a group given its perpendicular distance *y* from the transect line, and $\underline{c}$, a vector with its size and additional covariates. Since transect lines are allocated randomly in the survey area, perpendicular distances are assumed independent of group size and other covariates. With Y assumed uniform $I(0, W)$, the underlying joint distribution function $f(y,\underline{c})$ is

$$f(y,\underline{c}) = f(y)\,f(\underline{c}) = W^{-1} f(\underline{c}),$$

and the probability that a group is detected and has size and additional covariates $\underline{c}$ is

$$g(y,\underline{c})W^{-1} f(\underline{c}).$$

Similar to the univariate case, the unconditional probability of detecting a group is

$$P = \iint_W g(y,\underline{c}) f(y,\underline{c}) dy d\underline{c} = W^{-1} \iint_W g(y,\underline{c}) f(\underline{c}) dy d\underline{c},$$

so that the joint *pdf* of observed perpendicular distances and group size and additional covariates is

$$f(y,\underline{c}) = \frac{g(y,\underline{c})\, f(\underline{c})}{\iint_W g(y,\underline{c}) f(\underline{c}) dy d\underline{c}}.$$

This function, similarly derived by Drummer and McDonald (1987) and Quang (1993) for the bivariate case ($\underline{c} = s$), and by Borchers et al. (1998) and Chen (1999) for multiple covariates, requires the assumption of a certain form for $f(\underline{c})$. Thus, it is better to use a conditional *pdf* of y given $\underline{c}$,

$$f(y \mid \underline{c}) = \frac{g(y,\underline{c})}{\int_0^W g(y,\underline{c}) dy} = \frac{g(y,\underline{c})}{\mu(\underline{c})}, \tag{8}$$

justified by the theory of weighted distributions (Patil and Ord, 1976), where $\mu(\underline{c})$ is the effective strip half-width given covariates $\underline{c}$.

Assuming perfect detection on the transect line, the conditional *pdf* can be evaluated at 0 to estimate $\mu(\underline{c})$

$$f(0 \mid \underline{c}) = \frac{g(0 \mid \underline{c})}{\mu(\underline{c})} = \mu(\underline{c})^{-1}. \tag{9}$$

A mean estimate of $\hat{f}(0)$ is obtained as

$$\hat{f}(0) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\mu_{c_i}} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i), \tag{10}$$

so that, from Eq. *1*, density of groups can be estimated as

$$\hat{D} = \frac{\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i)}{2L}. \tag{11}$$

Density of individuals can be derived from Eqs. *2* and *10*,

$$\hat{D}_{ind} = \frac{\hat{E}(s)\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i)}{2L}, \tag{12}$$

and $\hat{E}(s)$ can be estimated as in Eqs. *5* or *7*. Otherwise, using a moments approximation of $\hat{E}(s)$ in Eq. *2*,

$$\hat{D}_{ind} = \frac{n\,\hat{f}(0)\frac{\sum_{i=1}^{n} s_i}{n}}{2L} = \frac{\sum_{i=1}^{n} \hat{f}(0) s_i}{2L}.$$

Under the multivariate case there is an independent $\mu(\underline{c}_i)$ for each detection, given its size and additional covariate values, and it can be shown that

$$\hat{D}_{ind} = \frac{\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i) s_i}{2L}. \tag{13}$$

Abundance of individuals is then estimated as $\hat{N} = A\hat{D}$,

$$\hat{N} = A \frac{\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i) s_i}{2L} \tag{14}$$

where A is the size of the surveyed area. Equations *13* and *14* are justified assuming a Horwitz-Thompson Line Transect estimator (Borchers et al., 1998; Marques, 2001).

From Eqs.*11* and *13*, an unbiased estimate of $\hat{E}(s)$ is obtained as a weighted average of the observed group sizes s_i, using the estimated effective strip half-width at each sighting as weight

$$\hat{E}(s) = \frac{\hat{D}_{ind}}{\hat{D}} = \frac{\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i) s_i}{\sum_{i=1}^{n} \hat{f}(0 \mid \underline{c}_i)}. \tag{15}$$

Parametric estimation

Group size and additional covariates can be included in the detection function as part of the scale and/or the shape parameter. In practice, the most efficient approach is through the scale parameter only, as in the bivariate exponential models of Drummer and McDonald (1987) and the multivariate exponential power series of Ramsey et al. (1987). Palka (1993) proposed a similar bivariate log-linear modeling of the scale parameter for the hazard rate function of Hayes and Buckland (1983), and here it is extended to account for multiple covariates in different models. The scale parameter is formulated as,

$$\varphi = e^{\alpha + \sum_{i=1}^{t} \alpha_i \log_e(c_i)}, \tag{16}$$

with covariates having multiplicative effects, or additive effects in the $\log_e$ scale. This parameterization allows for factors with several levels, continuous covariates with linear or quadratic effects such as group size, and interaction terms. φ was included in four detection functions: Pollock's (1978) exponential power series

$$g(y, \underline{c}) = e^{-\left(\frac{y}{\varphi}\right)^b},$$

with the half-normal (Quinn and Gallucci, 1980) $g(y,\underline{c}) = e^{-\frac{1}{2}\left(\frac{y}{\varphi}\right)^2}$ and the negative exponential model $g(y,\underline{c}) = e^{-\left(\frac{y}{\varphi}\right)}$ as special cases, and the hazard rate model (Hayes and Buckland, 1983)

$$g(y,\underline{c}) = 1 - e^{-\left(\frac{y}{\varphi}\right)^{-b}} .$$

In general, any suitable detection function including a scale parameter and meeting the "shape" criterion of Burnham et al. (1980) can be considered in this context. In this regard, the negative exponential will seldom be a good model.

A different approach, perhaps more flexible with sufficient sample size, is with semi-parametric models including expansion series (e.g. polynomial or cosine functions), to add further structure to the detection curves (Marques, 2001). In this paper I only explore simple parametric models.

Estimation and inference method. For a given parametric detection function $g(y,\underline{c})$, $\hat{f}(y\,|\,\underline{c})$ can be estimated with maximum likelihood methods

$$L\ (\theta; y, \underline{c}) = \prod_{i=1}^{n} \frac{g(y_i, \underline{c}_i)}{\int_0^W g(y, \underline{c}_i)dy},$$

where θ are the parameters of the detection function. Parameter estimates are obtained by maximizing the conditional likelihood function, using the Newton method or approximations, like the quasi-Newton algorithm used in the example analyses, and standard errors can be obtained from the inverse of the Hessian matrix $H(\hat{\theta})$.

From a series of candidate models, the most adequate parametric fit, including sequential selection of explanatory covariates can be based on the *AIC* (Akaike, 1973) or the modified AIC_c for small samples of Hurvich and Tsai (1995),

$$AIC_c = -2\log\{L\ (\theta; y, \underline{c})\} + 2k + \frac{k(k+1)}{n-k-1},$$

where *k* is the number of parameters and *n* is the sample size. The simple *AIC* tends to favor the overfit, whereas the AIC_c penalizes an increased number of parameters relative to sample size, and performs better with small and non-small sample sizes (Burnham and Anderson, 1998).

A suite of candidate models may have similar good abilities to describe the data, and there is no need to assume that a single model is superior. A flexible approach to account for model selection uncertainty is model averaging (Buckland et al., 1997) based on *AIC* weights. For *M* candidate models, the weight for model M_i is

$$AIC_c w_i = \frac{e^{-0.5\,\Delta AIC_c|M_i}}{\sum_{i=1}^{M} e^{-0.5\Delta AIC_{c_i}}},$$

and ΔAIC_c is the difference in AIC_c from a particular model (model M_i) with respect to the model with lowest AIC_c. An average estimate is obtained as

$$Ave\,\hat{f}(0\,|\,c) = \sum_{i=1}^{M} AIC_c w_i \hat{f}(0\,|\,c)_i \quad .$$

Nonparametric estimation

Conditional density functions can be estimated with kernel smoothers. I adapt Chen's (1999) global multivariate kernel smoother estimator to conventional line transect analysis. The conditional density function is

$$\hat{ff}(y\,|\,\underline{c}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-y_i}{h}\right) Z_i(\underline{c}), \tag{17}$$

where h is the smoothing bandwidth, K is a kernel function (e.g., a Gaussian kernel, $K(y) = \sqrt{2\pi}^{-1} e^{-y^2/2}$), and $Z_i(\underline{c})$ is the weight contributed by the covariates at each observed detection $(y_i, \underline{c}_i)$,

$$Z_i(\overline{c}) = \frac{\prod_{j=1}^{m} K\left(\frac{c_j - c_{ij}}{b_j}\right)}{\sum_{i=1}^{n}\prod_{j=1}^{m} K\left(\frac{c_j - c_{ij}}{b_j}\right)}, \tag{18}$$

where the same kernel is used for each of the m covariates, but each covariate has a different smoothing bandwidth, b_m. Ideally, signed y values (i.e., the sign indicates the side of the transect line) should be used to avoid *end effects* (Silverman, 1986) at perpendicular distance 0. Otherwise, distances and additional covariates can be reflected and sampled about an endpoint t. Density is then computed for the sample (y, $2t$-y) and doubled at $[t,W]$, or at $[t,\infty)$, because kernel smoothers do not necessarily require a truncation distance. Since K is symmetrical about t=0, the $\mu(\underline{c}_i)^{-1}$ can be estimated as

$$\hat{ff}(0\,|\,\underline{c}_i) = \frac{2}{nh}\sum_{i=1}^{n} K\left(\frac{y_i}{h}\right) Z_i(\underline{c}), \tag{19}$$

reflecting the covariate values also. That is, for each covariate c_j, that data becomes

$$K\left(\frac{c_j - c_{ij}}{b_j}\right) + K\left(\frac{c_j + c_{ij}}{b_j}\right).$$

Optimal bandwidths can be obtained minimizing the mean integrated squared error (MISE) of $\hat{f}$,

$$MISE(\hat{f}, h, b_m) = \int E\left\{\hat{f}(y,\underline{c}) - f(y,\underline{c})\right\}^2 dyd\underline{c}\,.$$

Estimates of h and b_m minimizing the *MISE* are typically obtained by cross-validation or by reference to a standard distribution, usually $f(y\,|\,\underline{c}) \rightarrow N\left\{0, \sigma^2(\underline{c})\right\}$.

The normal approximation is less computationally involved and performs well if the assumed distribution is close to the underlying distribution. Thus, a simple solution is to use a multidimensional approach (Scott, 1992) to estimate the smoothing parameters $h_i = \sigma_i \left\{4^{-1}(d+2)n\right\}^{(d+4)^{-1}}$, where d is the number of dimensions and σ_i the standard deviation in dimension i, which is replaced by a sample estimate for practical implementation.

VARIANCE AND CONFIDENCE INTERVAL ESTIMATION

Analytical variances of $\hat{D}_{ind}$ and $\hat{N}$ using univariate or multivariate detection functions can be approximated with the Delta method (e.g., Seber, 1982). For the multivariate case (Eq. *12*),

$$\hat{\text{var}}\left(\hat{N}\right) = A^2 \, \hat{\text{var}}\left(\hat{D}_{ind}\right) = A^2 D_{ind}^2 \left[\frac{\hat{\text{var}}\left\{\hat{f}(0 \mid \underline{c})\right\}}{\left\{\hat{f}(0 \mid \underline{c})\right\}^2} + \frac{\hat{\text{var}}\left\{\hat{E}(s)\right\}}{\left\{\hat{E}(s)\right\}^2} \right]$$

or (Eq. *14*)

$$\hat{\text{var}}\left(\hat{N}\right) = \left(\frac{A}{2L}\right)^2 \sum_{i=1}^{n} s_i^2 \, \hat{\text{var}}\left\{\hat{f}(0 \mid c_i)\right\}$$

However, knowledge of the covariates' underlying distribution function is required to obtain approximate conditional variances and covariances of $\hat{f}(0 \mid \underline{c})$.

A robust approach to estimate the variance and confidence interval, valid for both parametric and nonparametric detection functions, is the nonparametric bootstrap resampling from multiple transects, provided that transects are independent (Buckland et al., 2001). In each of B bootstrap replicates, transects with their corresponding detected groups are resampled with replacement until the total line length is equal or approximately equal to the total surveyed length. Otherwise, a balanced bootstrap (Davison and Hinkley, 1997) can be used, in which each transect line is selected as many times as bootstrap replicates. The parameter of interest is then estimated at each replicate, and the variance is estimated as the sample variance of the bootstrap replicates. Approximate confidence intervals can be obtained with the percentile method, or derived alternative methods, like the *BCa* confidence intervals (Efron and Tibshirani, 1993). The *BCa* confidence intervals have better properties than the simple percentiles and produce more realistic confidence limits if data sets are not small.

With parametric detection functions, the bootstrap allows for model selection uncertainty if, for example, selection of the model and explanatory covariates and model averaging based on AIC_c is carried out at each replicate. Another advantage of the bootstrap is allowing for model uncertainty when using regression methods to estimate $\hat{E}(s)$. At each bootstrap replicate, tests for bias of the final *M* and *LS* regression estimates, selection of covariates explaining $\hat{E}(s)$ by *RFPE*, and significant regression slope can be carried out, following the procedure proposed in the robust-regression section.

Modeling group size distribution

The observed group size distributions of many animal populations are highly skewed. In extreme cases, the distribution's right tail is very long and discrete because of rarely encountered very large groups (Fig.1). These large groups significantly influence the shape of the group size distribution and its mean; missing just one very large group implies a substantial negative bias in abundance and its variance. Thus, simple bootstrap resampling of the observed values may fail to produce confidence intervals with optimum coverage and variances are usually underestimated. Improved bootstrap resampling can be obtained modeling the group size distribution and resample from it at each bootstrap replicate.

Parametric models or combinations of these (e.g., mixture models) can be used to model group size distributions. In general, a mixture of one or more Gamma or lognormal distributions for the lower group sizes and an extreme value distribution for the right tail will be reasonable options. A simple mixture of a lognormal and Pareto distributions is

$$f(s \mid \pi, \mu, \sigma, \psi) = \pi_1 \left[\frac{1}{s\sigma} e^{-0.5\left\{(\log_e s - \mu)\sigma^{-1}\right\}^2} \right] + (1 - \pi_1)\left(\frac{\psi}{s^{\psi+1}} \right),$$

with ψ as the shape parameter of the Pareto distribution, π is the proportion of the distribution to which the first parametric model (e.g. lognormal) is fitted, and s $\geq$ 1.

Alternatively, group size distributions can be modeled with nonparametric methods that are more data oriented. A flexible approach (as used in the example analysis) is with adaptive kernel smoother density estimation

$$\hat{ff}(s) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i h} K\left(\frac{s - s_i}{\lambda_i h} \right), \tag{20}$$

with a Gaussian kernel *K* and local smoothing weights λ_i. The weights select larger bandwidths in low-density areas and smaller ones in high-density areas, better modeling the distribution's right tails.

Following Silverman (1986), smoothing weights can be estimated as $\lambda_i = \left\{ f_p(s_i) G^{-1} \right\}^{-1/2}$, where $f_p(s_i)$ is a pilot density estimate, and *G* is the geometrical mean of $f_p(s_i)$. A reliable bandwidth *h* can be obtained with a *plug-in* method (Sheather and Jones, 1991) minimizing the asymptotic *MISE*. In this method, an asymptotically optimal h_0 (e.g. from a normal approximation) is initially used to obtain $\hat{h}$ as

$$\hat{h} = \left\{ \frac{R(K)}{\sigma_K^4 \hat{R}(f'')} \right\}^{1/5} n^{-1/5},$$

where $R(\rho)$ is $\int \rho(u)^2 du$, and substituting $\hat{R}(f'')$ by $R(\hat{f}'')$. The *f''* is replaced by a kernel estimate from the data.

A less computationally involved option is non-adaptive kernel density estimation,

$$\hat{ff}(s) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{s - s_i}{h}\right),$$

with $h = \sigma\left\{4(3n)^{-1}\right\}^{1/5}$, and Hogg's (1979) median absolute deviation estimator $\hat{\sigma} = median\left\{\left|s_i - \hat{v}\right|\right\}0.6745^{-1}$, where $\hat{v}$ is the median of the sample of group sizes. This estimator of *h* takes into account the effects of the long tail distribution.

Resampling procedure for group size

At each bootstrap replicate, transect lines are selected with replacement, and the resampled group sizes are used as in a shrunk smoothed bootstrap (Davison and Hinkley, 1997), generating new group size values from the cumulative distribution function obtained with Eq.20.

For practical implementation, group size values are generated as $s_i^* = s_{I_i^*} + \lambda_i h \varepsilon_i$, for i = 1,...*n*, where the I_i^* are independent and uniformly distributed on the integers 1,...,*n*, which in this case are a sample from independent transect lines. The ε_i are a random sample from $K\left\{\left(s - s_i\right)\left(\lambda_i h\right)^{-1}\right\}$, independent of the I_i^* (i.e. random deviates from a standard normal distribution). The s_i^* are rescaled by its variance, $n^{-1}\sum_{i=1}^{n}\left(s_i - \overline{s}\right)^2 + (\lambda_i h)^2$, to have the same variance as the unsmoothed distribution. With the rescaled group sizes and the associated covariate data parameters of interest are computed at each replicate, and the variance and confidence interval are obtained as explained above.

EXAMPLE ANALYSIS

The proposed methods are applied to line transect data of the eastern spinner dolphin (*Stenella longirostris orientalis*) stock in the ETP collected during research cruises in 1998, 1999 and 2000. Methods of data collection followed standard protocols for line-transect surveys conducted by the Southwest Fisheries Science Center (Wade and Gerrodette, 1993; Kinzey et al., 2000). Searching effort was randomly allocated and stratified into four areas (Fig. 2): a core stratum (5,869,485 km^2) centered on the main stocks of interest, an outer stratum (14,777,855 km^2), a north coastal stratum (534,821 km^2), and a south coastal stratum (171,464 km^2). Surveys were carried out from late July to early December with the oceanographic ships *RV Endeavor* from the University of Rhode Island and *RV David Starr Jordan* and *RV McArthur* from NOAA in 1998, and with the two NOAA ships only in 1999 and 2000. Perpendicular distances and additional covariates were collected at each sighting from the ship's flying bridge (over 10 m high), moving along the transect lines at a constant speed of 10 knots. Dolphin groups were primarily detected with pedestal-mounted 25x150 binoculars. Covariate data included total group size, species proportion in groups, detection cue (seabirds, water splashes, and animal body), association with seabirds (seabirds' presence/absence), sea state as measured by the Beaufort scale, swell height in meters, time of the day, visibility in km, a weather factor (rain, fog, etc.), glare, and ship. Given the large variability in group size (Fig. 1) and the corresponding observers' estimates, these were corrected for measurement error with calibration methods based on aerial photography (Barlow et al., 1998).

The proposed methods were compared using a sub sample corresponding to the combined effort and sightings data of the core and north coastal stratum for each year to estimate abundance and expected group size. Four analysis options were used: first, univariate parametric detection functions and expected group size estimated with the *LS* method, i.e. bias-correction was applied if the regression slope of $\log_e$ group size against detection probability was significantly different from 0 ($p<0.15$); second, univariate detection functions with robust-regression group size estimate (Eq.7), if the slope was significant and tests indicated no bias of the final *M*-estimates; third, parametric multivariate analysis of the detection function in Eqs. *14* and *15*; and fourth, nonparametric multivariate analysis of the detection function in Eqs. *14* and *15*.

Model selection and averaging was carried out in the parametric analysis. With nonparametric estimation there is not an obvious objective model selection method and thus, covariate effects were explored in a general way. In the simplest case of two covariates, perpendicular distance and group size, I used a likelihood ratio expression (Bowman and Azzalini, 1997) of the joint and marginal density functions,

$$LRT = \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\hat{f}\left(y_i,c_{1i}\right)}{\hat{f}\left(y_i\right)\hat{f}\left(c_{1i}\right)}\right\}, \tag{21}$$

where c_1 is group size or another covariate. The distribution of this test statistic was approximated by permutation. Permuted values of c_{1i} were randomly associated to values of y_i and, under the null hypothesis of independence of covariates, an empirical p-value was computed from the proportion of permuted statistic values higher than the observed value.

Further testing was done by comparing bivariate distributions among groups with the statistic

$$\sum_{i=1}^{r} m_i \iint\left\{\hat{f}_i(y_i,c_{1i}) - \hat{f}(y_i,c_{1i})\right\}^2 dy dc_1\,, \tag{22}$$

where the $\hat{f}_i\left(y_i,c_{1i}\right)$ were the estimated densities for the different *m* groups, and $\hat{f}\left(y_i,c_{1i}\right)$ was the estimated joint density across groups. In the example dataset, bivariate density functions of perpendicular distance and group size between levels of different factors were compared. The statistic was constructed by numerical integration, using a two-dimensional grid of evaluation points, and comparisons were based on the bivariate densities by groups. For example, with seabird association, the null hypothesis was of no differences between bivariate densities of perpendicular distance and group size with and without seabirds. Sightings with seabirds were sampled with replacement in an ordinary bootstrap procedure, and the resampling indices were used to obtain bootstrap data sets for sightings without seabirds. The empirical P-value was computed as indicated in the previous test, and the number of bootstrap replicates was 1000.

Balanced bootstrap resampling was used with modeling of the group size distribution. The modeling of school size was with adaptive kernel smoothers (Eq. *20*), with optimal bandwidths estimated with the *plug-in* Sheather-Jones method. In the parametric multivariate analysis, model selection and averaging was also carried out at each replicate to estimate the variance. In the nonparametric analysis, testing covariate effects (Eqs. *21*, *22*) at each bootstrap replicate was extremely inefficient and it was only carried out for an initial selection of potential covariates. The covariates with more significant effects were all modeled at each replicate. All bootstrap confidence intervals were estimated with the *BCa* method.

Results

Of the four methods tested, the analysis with parametric multivariate detection functions and the analysis with robust regression provided more consistent abundance estimates over time. The nonparametric analysis and the *LS* bias-correction analysis underestimated the mean group size and produced lower abundance estimates. The fit of the two regression methods to the example data sets is illustrated in Figures 3a-c.

In 1998, the *LS* fit was significant (P=0.0408) according to the standard significance level used with this method, i.e., $P < 0.15$ (Thomas et al. 1998). The robust test, however, indicated that the *LS* estimate was significantly biased (P=0.0167), and most of the bias was caused by a very small group detected on the transect line (Fig. 3a, arrow). In contrast, the robust *MM* fit was not biased (P=0.8314). With it, the estimated slope was also significant (P=0.0819), but a lower correction was applied to the mean group size (Table 2). In 1999, the *LS* fit was also significant (P=0.0026) and found to be biased by the robust test (p=0.0082). The bias was caused again by an outlier (arrow in Fig. 3b). The *MM* regression fit was not biased (P=0.9563), and the slope was also significant (P=0.0056). Thus, a different mean group size correction was applied with each method (Table 2). In year 2000, neither regression fit was biased (*LS*: P=1; *MM*: P=0.4424), slopes were significant with the *LS* (P=0.1166) and non-significant with the *MM* method (P=0.3401), and the mean group size was only corrected with the *LS* analysis (Table 2).

Looking into the bootstrap replicates of all years, the *LS* method often produced significant regression slopes of log-group size against $g(y)$, leading to low mean group size estimates. The same slopes were lower or non-significant with the *MM* regression, which imposed a lower or no correction at all when the association between total group size and perpendicular distance was highly nonlinear. As a result, variance estimates were high, with perhaps a slight positive bias, but were less biased than those obtained with the *LS* method.

The association between perpendicular distance and the $\log_e$ group size was further investigated with nonparametric bivariate density functions. The probability density contours of the 1998 data (Fig. 3d) did not indicate a clear linear association between variables, given the high aggregation of observations with average group size in the first 2 km off the transect line. In contrast, a likelihood ratio test (LRT Eq. *21*, 1000 permutations, empirical P<0.0001), indicated a definite correlation between perpendicular distance and group size. In agreement, the fit of a bivariate half-normal model with group size was better (AIC_c=280.96, n=99, Table 1a) than that of a univariate half-normal model (AIC_c = 284.96, Table 1a). Time of the day was also associated to perpendicular distance, as indicated by the LRT (P<0.0001), and the relatively good fit of the half-normal model with total group size and time of the day (ΔAIC_c = 2.611, Table 1a). With model averaging, the mean parametric $\hat{f}(0)$ was estimated as 0.3583 (%CV=8.33). The mean nonparametric $\hat{f}(0)$, modeled with total group size and time of the day, divided into four discrete categories because of low sample size, was estimated as 0.4271 (%CV=16.28).

In 1999, the bivariate density contours indicated a positive correlation between perpendicular distance and log-group size (LRT, P<0.0001, Fig. 3e). In agreement, the bivariate half-normal model with total group size fitted the data better (AIC_c=227.59; n=70, Table 1b) than the univariate half-normal (AIC_c=232.04, Table 1B). The hazard rate model with swell height (AIC_c=228.54, Table 1b) was the second best model. In agreement, the association of perpendicular distance and swell height was significant (LRT, P<0.0001). The bivariate density of perpendicular distance and group size was different for sightings with seabirds than without seabirds (Eq. *22* statistic = 0.2586; P<0.0125). Thus, seabirds was retained as a covariate for the nonparametric model, together with group size. In the parametric case, the hazard-rate model with seabirds was the third best model (AIC_c=230.25, Table 1b). With model averaging, mean $\hat{f}(0)$ was estimated as 0.2816 (%CV=11.54), and the nonparametric mean $\hat{f}(0)$, modeled with total group size and swell height, divided into four discrete categories, and seabirds was estimated as 0.2911 (%CV=20.69).

In year 2000, the bivariate density plot did not indicate a perpendicular distance effect on group size but there was a significant correlation (LRT, P<0.0001; Fig. 3f). Unlike in 1999, the bivariate densities were not different for sightings with and without seabirds (Eq. *22* statistic = 0.8429; P=0.5301). Differences in parametric detection function fit were small between the simple univariate half-normal model and bivariate half-normal models with swell height, total group size, and Beaufort (all ΔAIC_c< 2, n=70; Table 1c). With model averaging, the parametric $\hat{f}(0)$ was estimated as 0.3244 (%CV=13.79). The nonparametric $\hat{f}(0)$, modeled with total group size and swell height, divided into discrete categories because of the low sample size, was estimated as 0.3699 (%CV=22.86).

Nonparametric group size estimates were low and similar to the estimates obtained with the *LS* method (Table 2). Since these were driven by outliers and found to be biased by the robust test, it is likely that nonparametric estimates, with comparable values, were also negatively biased. In contrast, abundance estimates obtained with the two methods were quite different because of large differences in the $\hat{f}(0)$ estimates. The robust *MM*-regression method produced group size estimates similar to those obtained with the multivariate parametric models. Since the differences in the $\hat{f}(0)$ estimates produced with both methods were not large, abundance estimates were also similar (Table 2).

Abundance estimates were more consistent over time using multivariate parametric methods. With these methods they were also precise, at reverse of nonparametric estimates, which were only precise in year 2000. Estimates obtained with robust regression methods were comparable to those obtained with parametric multivariate detection functions but were less precise, given the higher variability in mean group size estimates. Estimates obtained with *LS* regression estimates of mean group size were more unstable, less consistent and biased.

In general, differences in effort and number of sightings made abundance estimates more precise in 1998. However, the numbers produced for Table 2 should only be used for comparison of the methods presented since they do not account for the entire distribution of the stock and can be misleading. Estimates for the complete Eastern spinner dolphin stock will be presented elsewhere.

SIMULATION TESTS

I present a summary of simulations designed to examine the proposed estimation methods in terms of bias and precision. Data were generated to recreate the high variability in group size and size-dependent detection observed in line transect surveys of ETP dolphin stocks. The observed sample size was obtained from a Poisson distribution with E(n) = 150, the truncation distance W was 5 km, and L was calculated so that density was 0.95 groups/km^2. The true distribution of group sizes was log-normal with mean μ=4.5 and σ^2=1, and constrained to range from 1 to 2000 dolphins, so that the true E(s) was 146 dolphins. For simplicity, I simulated only three covariates: perpendicular distance, group size and a factor with two levels (e.g., presence or absence of associated seabirds).

For a given group size, generated from a lognormal distribution, a level of seabird association was obtained from a binomial trial, and a perpendicular distance from a uniform distribution $I(0, W)$. For the detection function, a trivariate exponential power series was used:

$$g(y, c_1, c_2) = e^{-\frac{1}{2}\left(\frac{y}{e^{\alpha_1 \log_e c_1 + \alpha_2 \log_e c_2)}}\right)^b},$$

where y is perpendicular distance, c_1 is group size, and c_2 is presence or absence of seabirds, and the rejection method was used to select detected trials. Parameters α_1 and α_2 controlled the effects of group size and seabird association, and b is a shape parameter determining the width of a "shoulder" near the transect line.

Data were simulated for two different sets of parameters α_i, with b fixed at 1.5 to obtain a narrow left shoulder in the perpendicular distance histograms, as in the example data set. α_i was fixed at 0.2 imposing a moderate effect of group size on detection and α_2 was either 0.05 or 0.2, representing low and high effects respectively. The results are based on 1000 simulations and 400 bootstrap replicates, and the analysis options are the same as in the example analysis. Computed statistics include point estimates, standard errors, percent relative bias $PRB = 100(\hat{\theta} - \theta)/\theta$, mean square error $MSE = \hat{var}(\theta) + (\theta - \hat{\theta})^2$, where θ is the parameter of interest, and observed coverage of bootstrap confidence intervals with 95% nominal coverage.

Results

Group size estimates with the four methods were consistent with the true value. All were positively biased, with a slightly higher bias in those obtained with the parametric multivariate analysis and the robust regression analysis. The mean square error (MSE) was lower with the parametric multivariate analysis, which also produced more precise results. The nonparametric multivariate analysis had always the largest MSE (Table 3).

The *LS* analysis produced significant slopes 40 and 49% of the times with the two different sets of parameters respectively. The robust *MM*-regression analysis indicated that slopes were significant in 37 and 42% of the simulations. *LS* estimates were found to be biased 24 and 27% of the times, whereas bias in the final *M*-estimates only occurred 7 and 11% of the times. In 22 and 23% of the simulations the *LS* slopes were biased but the *M*-slopes were not, and only in 4 and 7% of the simulations the *M*-estimates were biased but the *LS* estimates were not. In 7 and 8% of the simulations the LS analysis was biased and indicated significant slopes whereas the *M*-estimates were non-biased and indicated significant slopes.

The mean $\hat{f}(0)$ estimates with univariate parametric functions, i.e. the estimates used with regression methods, had negative PRB, -3.3 and -0.4% for each set of parameters respectively, compared to the true $f(0)$ of 0.300. The parametric multivariate analysis produced the less biased, -1.9 and 0.93%, and also the more precise mean $\hat{f}(0)$ estimates. In contrast, the nonparametric mean $\hat{f}(0)$ estimates were positively biased, 2.2 and 5.8% respectively, were the less precise, and had the largest MSE.

Density estimates were less biased with multivariate methods when the effect of the third covariate, in addition to that of group size, was high; i.e. α_2=0.20. As expected, both regression methods produced more biased density estimates in that case (Table 3).

When the effect of the second covariate was not strong, regression methods produced less biased density estimates. In general, density estimates with multivariate parametric methods were more precise and had lower MSE, whereas the estimates with nonparametric methods were the less precise and had higher MSE.

Bootstrap confidence interval coverage was high for most simulations, near the nominal 95%, and was similar with the smoothed and non-smoothed group size methods. In general, the smoothed bootstrap tended to give a slightly better coverage, but in most cases differences in coverage were marginal. Variances estimated with the smoothed group size distribution were only slightly larger than those with the unsmoothed distribution. This indicates that the discreteness in the right tail of the distribution was better sampled with the smoothing algorithm. However, the simulations were not designed to capture the *edge effect* of the right tail; i.e. group sizes were constrained to a maximum value of 2000. In agreement, no values substantially larger than 2000 were found upon examination of the smoothed bootstrap distributions.

DISCUSSION

Linear regression using least squares to model group size can be very sensitive to the effects of just a single observation (Fig. 3a,b) and produce biased estimates. Besides, using a $\log_e$ transformation not always ensures that group size associates linearly with detection probability, perpendicular distance or other covariates. Thus, imposing a linear correction with highly variable group sizes, low sample size and outliers may not always be adequate; abundance estimates will be biased and variable. Robust regression is also constrained by the linearity in the relationship between covariates but it accounts effectively for outlier effects and will produce better results than the *LS* regression.

Robust regression also depends on the probability level imposed to test for a statistically significant slope. It can be argued, as it is done to support the use of the *LS* method (Thomas et al., 1998), that the linear modeling will efficiently correct group size bias regardless of finding a significant slope. In the light of the results of this paper, this practice also requires the testing for bias in the outcomes of whatever linear modeling method of choice. The methods I propose are an example of how this can be easily achieved.

The covariate analysis of the detection function successfully reduces heterogeneity and it should be considered *a priori*, even if only for testing covariate effects like size dependent detection. In most cases, the parametric analysis is likely to improve density and abundance estimates with an adequate model selection and reasonable sample sizes. It is always to be preferred to stratification or post-stratification of the data because it models heterogeneity directly. For instance, it is very useful to test for geographic strata, year or species when these effects are modeled as covariates. Model selection provides an objective statistical criterion to combine strata for estimation when precision is to be gained from pooling.

A down side of considering multiple covariates is multi-model inference with a very large set of good candidate models. This is the case of the example data set for year 2000, with 17 models with ΔAIC_c smaller than 4. Since the purpose of the modeling is inference, it may be justified to use them all in model averaging because they have substantial or good support (Burnham and Anderson, 2001). However, the amount of computing, especially in the bootstrap resampling, is perhaps too high to justify using all the models and a sensitivity analysis can help deciding whether considering just models with $\Delta AIC_c \leq 2$ is a good alternative. Here is also where the science of the problem should be considered. For instance, if a time series is to be analyzed and further used in a population modeling, like the example data set, it is arguable that the same family of parametric detection function models should be used for each year for consistency. Further, in cases like the example data set, research on the problem of why the data is spiked may discard or justify the use of models that overfit, such as the hazard rate.

The covariate analysis requires large sample sizes to obtain reliable results. Multivariate models will seldom fit well sparse and small datasets unless some particular covariate effect is high. This problem will be exacerbated with nonparametric methods, where every covariate is an additional dimension to the analysis. As noted by Silverman (1986), as the number of dimensions increases, accuracy and precision decrease exponentially. An alternative nonparametric method could be Chen's (1999) local kernel smoother estimator modified for the conventional case. Its advantage is that it is univariate, unlike the method used in this paper, and perhaps is more efficient. However, it also requires discretization of continuous covariates and this might not be an optimal practice when sample size is low. Moreover, there is not an obvious selection process to eliminate covariates with marginal effects. Ultimately, it is always better to conduct pilot surveys and address sample size issues during the survey design, so that the analysis can better accommodate the needs of a particular data set.

An additional common problem with nonparametric estimation is the lack of smoothness of perpendicular distances near the transect line. This is also a problem with kernel smoothers with adaptive bandwidth selection (Chen, 1996), and it was found in the simulations and also in the example data. In the simulations, data sets were deliberately spiked to mimic the example data, and it was seen how kernel smoothers overestimated the mean $\hat{f}(0)$. In the example analysis, however, it is not clear whether the spike was purely a sampling artifact or if the true detection function was in fact spiked. In any case, data oriented models like the hazard-rate or kernel smoothers tended to fit the spike and produce higher estimates of $\hat{f}(0)$. As noted by Buckland et al. (2001) models like the hazard-rate produce positively biased $\hat{f}(0)$ estimates in these cases. Thus, a family of more flexible detection functions can be used as potential starting candidates. Some assessment of the fit near 0 and of meeting the shape criterion (Burnham et al., 1980) by models will be helpful. Good candidate models can be simple parametric ones as proposed in this paper, their extension with expansion series (Marques, 2001), or the extension with other semi-parametric approaches like in the univariate case (Barabesi, 2000). On the other hand, kernel smoothers have good properties, like the ability of dealing better with local nonlinearities. Adaptive kernel smoothers could perhaps be ameliorated with better methods of bandwidth selection (Barabesi and Fattorini, 1994, Gerard and Schucany,1999) and, with an objective covariate selection method, become good analysis tools.

The modeling of group size slightly improved the coverage of the confidence intervals. It is arguable that such computationally involved method should be used on a regular basis and it may not be justified unless very large schools occur. However, any analysis will be problematic with highly variable group sizes as in the example data set. Thus, modeling the group size distribution is a good tool for a sensitivity analysis and better assess the effect of very large groups. In particular, the modeled distributions can be used to assess the extent of the right tails and assess if these are representative of the observed data when resampling.

The examination of the distribution tails can also be helpful when fitting parametric detection function models. Differences in abundance can be substantial by including or excluding just a single large group size. This happens, for example, by right-truncation of perpendicular distances. Right-truncation is especially recommended when using Horvitz-Thompson-like estimators, as in Eq. *14*, because the $\hat{f}(0\,|\,\underline{c})$ estimates enter the estimator as part of the inclusion probabilities, and may bias the results if these are too small (Borchers et al., 1998; Marques, 2001). It is also true that if detection is highly size-biased, abundance estimates may be more sensitive to truncation of very large groups, usually detected at very large distances. In some cases, even a very large group on the right edge of the selected truncation distance can make a difference. Thus, truncation is a trade-off of good inclusion probabilities and good representativeness of large group sizes.

Dealing with highly variable group sizes, the *LS* regression analysis of mean group size and the multivariate nonparametric methods should be used with caution. Both methods can easily produce biased estimates. In the nonparametric analysis, the lack of good fit to short perpendicular distances is the main cause of bias, and further development is needed to take advantage of the good aspects of the method. In the case of the *LS* regression, the sensitivity to outliers is of major concern. Even if the method can work well, as seen in the simulations, it is likely that the simulated data did not recreate the severity of outlier effects, as in the example data set, and tests for bias should be common practice when analyzing real data. Robust regression techniques can be very helpful to test for such effects and, if necessary, can be used for inference. However, robust regression will tend to produce less precise abundance estimates, given the high variability of mean group size estimates using this method. In addition, the estimates can be biased if the effect of a covariate other than group size is strong. Thus, parametric multivariate methods, which tend to produce more precise estimates, will provide the best trade-off of bias and precision.

ACKNOWLEDGMENTS

This work was developed as part of the requirements of the International Dolphin Conservation Program Act (IDCPA, 16 U.S. Code 1414). The research was conducted thanks to the author's National Research Council research associateship at the Sowthwest Fisheries Science Center. Tim Gerrodette, Jay Barlow, Sharon Hedley, Paul Medley and Robert Mohn reviewed early versions of this manuscript and provided very useful comments. Tim Gerrodette, Sharon Hedley, John Brandon, Katie Kramer, Doug Kinzey, Paula Olson, Al Jackson, Steve Buckland, Cleredy Lennert-Cody, Tore Schweder, Robert Holland, Tom Hedley, and Karl Strauss contributed comments, ideas, criticism, support and/or logistics while the author was developing this work.

LITERATURE CITED

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In International Symposium on Information Theory, 2nd edition (eds. B. N. Petran and F. Csàaki), Acadèemiai Kiadi, Budapest, Hungary, pp. 267-81.

Barabesi, L. and L. Fattorini. 1994. A note on bandwidth selections for kernel density estimators at endpoints. Metron 52:43-56

Barabesi, L. 2000. Local likelihood density estimation in line transect sampling. Environmetrics 11:413-22.

Barlow, J., T. Gerrodette and W. Perryman. 1998. Calibrating group size estimates for cetaceans seen on ship surveys. Administrative Report No. LJ-98-11, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037. 39 pp.

Borchers, D. L., S. T. Buckland, P. W. Goedhart, E. D. Clarke and S. L. Hedley. 1998. Horvitz-Thompson estimators for double-platform line transect surveys. Biometrics 54:1221-1237.

Bowman, A. W. and A. Azzalini. 1997. Applied smoothing techniques for data analysis. Clarendon Press, Oxford, UK.

Buckland, S. T., K. P. Burnham and N. H. Augustin. 1997. Model selection: an integral part of inference. Biometrics 53:603-618.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J.L. Laake, D. L. Borchers and L. Thomas. 2001. Introduction to distance sampling. Estimating abundance of biological populations. Oxford University Press, Oxford.

Burnham, K. P. and D. R. Anderson. 1976. Mathematical models for nonparametric inferences from line transect data. Biometrics 32:325-336.

Burnham, K. P., D. R. Anderson and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. Wildlife Monographs 72:1-203.

Burnham, K. P. and D. R. Anderson. 1998. Model selection and inference. A practical information-theoretic approach. Springer-Verlag, New York, USA.

Burnham, K. P. and D. R. Anderson. 2001. Kullback-Leibler information as a basis for strong inference in ecological studies. Wildlife Research 28:111-119.

Chen, S. X. 1996. Studying school size effects in line transect sampling using the kernel method. Biometrics 52:1283-1294.

Chen, S. X. 1999. Estimation in independent observer line transect surveys for clustered populations. Biometrics 55:754-759.

Davison, A. C. and Hinkley, D. V. 1997. Bootstrap methods and their application. Cambridge university Press, Cambridge, UK.

Drummer, T. D. and L. L. McDonald. 1987. Size bias in line transect sampling. Biometrics 43:13-21.

Efron, B. and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and hall, London.

Gerard, P. D. and W. R. Schucany. 1999. Local bandwidth selection for kernel estimation of population densities with line transect sampling. Biometrics 55:769-773.

Hayes, R. J. and S. T. Buckland. 1983. Radial-distance models for the line-transect method. Biometrics 39:29-42.

Hogg, R. V. 1979. Statistical robustness: one view of its use in applications today. American Statistician 33:108-116.

Hurvich, C. M. and C. L. Tsai. 1995. Model selection for extended quasi-likelihood models in small samples. Biometrics 51:1077-84.

Kinzey, D., P. Olson and T. Gerrodette. 2000. Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. Administrative Report No. LJ-00-08, NMFS, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92037. 32 pp.

Marques, F. F. C. Incorporating covariates into standard line transect analyses. *In*: Estimating wildlife distribution and abundance from line transect surveys conducted from platforms of opportunity. Unpublished Ph.D. thesis, University of St. Andrews, St. Andrews, Scotland, May 2001.

MathSoft Inc. 1999. S-PLUS 2000 Guide to Statistics, Volume 1, Data Analysis. Products Division, MathSoft, Seattle, WA.

Palka, D. L. 1993. Estimating density of animals when assumptions of line-transect surveys are violated. Unpublished PhD. thesis, University of California, San Diego, USA.

Patil, G. P. and J. K. Ord. 1976. On size-biased sampling and related form-invariant weighted distributions. Sankyã, Series A 38:48-61.

Perrin, W. F. 1969. Using porpoises to catch tuna. World Fishing 18:42-45.

Pollock, K. H. 1978. A family of density estimators for line-transect sampling. Biometrics 34:475-478.

Quang, P. X. 1991. A nonparametric approach to size-biased line transect sampling. Biometrics 47:269-279.

Quinn, T. J. and Gallucci, V. F. 1980. Parametric models for line transect estimators of abundance. Ecology 61:292-302.

Ramsey, F. L., V. Wildman and J. Engbring. 1987. Covariate adjustments to effective area in variable-area wildlife surveys. Biometrics 43:1-11.

Rousseeuw, P. J. and V. Yohai. 1984. Robust regression by means of S-estimators. In: Robust and nonlinear time series analysis, J. Franke, W. Hardle, and R. D. Martin (eds.), Lecture notes in statistics 26:256-272, Springer-Verlag, New York.

Scott, D. W. 1992. Multivariate density estimation. Wiley, New York, USA.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Charles Griffin and Co. Ltd, London.

Sheather, S. J. and M. C. Jones. 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B, 53:683-690.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, London.

Thomas, L., J.L. Laake, J. F. Derry, S.T. Buckland, D. L.Borchers, D. R. Anderson, K. P. Burnham, S. Strindberg, S. L. Hedley, M. L. Burt, F. Marques, J. H. Pollard and R. M. Fewster. 1998. Distance 3.5. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK.

Wade, P. R. and T. Gerrodette. 1993. Estimates of cetacean abundance and distribution in the eastern tropical Pacific. Reports of the International Whaling Commission 43:477-493.

Yohai, V., W. A. Stahel and R. H. Zamar. 1991. A procedure for robust estimation and inference in linear regression. In W. A. Stahel and S. W. Weisberg (eds.), Directions in robust statistics and diagnostics, Part II, Springer-Verlag, New York.

Table 1a-c. Parametric detection function analysis of perpendicular distance (*pd*) and additional covariate data for 1998, 1999, and 2000. Models considered are the half-normal, the exponential power series and the hazard rate. Potential covariates (see text for details) are total group size (*gs*), time of the day (*time*), swell height (*sh*), sea state measured by Beaufort scale (*Bf*), association with birds (*bi*), sighting cue (*cue*) and ship (*ship*). Model selection is based on the AIC_c statistic. ΔAIC_c is the difference between the a model's AIC_c and the lowest AIC_c. Estimates of $\hat{f}(0)$ are provided for comparison among models, and average estimates are provided in Table 2. AIC_cw are the weights of each model used in the model averaging.

Table 1a. Eastern spinners dolphins in 1998.

Model	AIC_c	ΔAIC_c	AIC_c**w**	$\hat{f}(0)$
Half-normal (pd+gs)	280.96	0	0.542	0.3584
Half-normal (pd+gs+time)	283.36	2.611	0.147	0.3689
Half-normal (pd)	284.96	4.004	0.073	0.3544
Exponential-power (pd)	285.23	4.274	0.064	0.4046
Hazard-rate (pd)	285.85	4.894	0.047	0.4176
Exponential-power (pd+gs)	285.91	4.951	0.045	0.3122
Hazard-rate (pd+time)	288.08	7.122	0.015	0.4156
Half-normal (pd+sh)	288.13	7.174	0.015	0.3429
Exponential-power (pd+time)	288.31	7.350	0.014	0.3683
Hazard-rate (pd+bi)	289.27	8.310	0.008	0.3348
Half-normal (pd+gs+sh)	289.88	8.903	0.006	0.3477
Hazard-rate (pd+gs)	291.24	10.279	0.003	0.2821
Exponential-power (pd+bi)	291.53	10.570	0.003	0.3099

Table 1b. Eastern spinners dolphins in 1999.

Model	AIC_c	ΔAIC_c	AIC_c**w**	$\hat{f}(0)$
Half-normal (pd+gs)	227.59	0	0.377	0.2861
Hazard-rate (pd+sh)	228.54	0.954	0.234	0.2013
Hazard-rate (pd+bi)	230.25	2.665	0.099	0.2015
Exponential-power (pd)	230.33	2.746	0.095	0.2049
Hazard-rate (pd+time)	231.26	3.669	0.060	0.2065
Half-normal (pd)	232.04	4.456	0.041	0.2491
Half-normal (pd+ship)	232.50	4.950	0.032	0.2640
Half-normal (pd+sh)	233.96	6.376	0.015	0.2557
Half-normal (pd+time)	234.12	6.535	0.014	0.2488
Hazard-rate (pd)	235.64	8.056	0.007	0.2753
Half-normal (pd+bi)	235.65	8.060	0.007	0.2664
Half-normal (pd+bf)	236.14	8.552	0.005	0.2650
Half-normal (pd+time+ship)	236.88	9.291	0.004	0.2803

Table 1c. Eastern spinners dolphins in 2000.

Model	AIC_c	ΔAIC_c	AIC_c**w**	$\hat{f}(0)$
Half-normal (pd)	214.99	0	0.142	0.3159
Half-normal (pd+sh)	215.60	0.606	0.105	0.3175
Half-normal (pd+gs)	216.20	1.202	0.078	0.3187
Half-normal (pd+bf)	216.77	1.778	0.058	0.3241
Exponential-power (pd)	216.94	1.932	0.054	0.2952
Half-normal (pd+time)	217.05	2.059	0.051	0.3163
Half-normal (pd+ship)	217.31	2.312	0.045	0.2998
Half-normal (pd+sh+bf)	217.39	2.399	0.043	0.3364
Half-normal (pd+bi)	217.55	2.553	0.040	0.2938
Exponential-power (pd+sh)	217.73	2.733	0.036	0.2773
Hazard-rate (pd)	217.74	2.750	0.034	0.2544
Half-normal (pd+sh+time)	217.79	2.795	0.035	0.3229
Hazard-rate (pd+sh)	217.87	2.876	0.034	0.2519
Hazard-rate (pd+gs)	218.24	3.246	0.028	0.2577
Exponential-power (pd+gs)	218.25	3.257	0.028	0.2993
Half-normal (pd+sh+gs)	218.35	3.358	0.026	0.3159
Exponential-power (pd+time)	218.99	3.997	0.019	0.2910
Exponential-power (pd+bf)	219.22	4.228	0.017	0.3051
Half-normal (pd+cue)	219.32	4.324	0.016	0.3061
Exponential-power (pd+ship)	219.39	4.399	0.016	0.2999
Half-normal (pd+sh+ship)	219.43	4.437	0.015	0.2979
Half-normal (pd+sh+bi)	219.48	4.483	0.015	0.2967
Exponential-power (pd+bi)	219.64	4.650	0.014	0.2938
Hazard-rate (pd+time)	219.76	4.762	0.013	0.2574
Hazard-rate (pd+ship)	220.02	5.024	0.011	0.2828
Hazard-rate (pd+bf)	220.50	5.508	0.009	0.2974
Hazard-rate (pd+bi)	220.53	5.539	0.009	0.2729
Half-normal (pd+sh+cue)	221.90	6.904	0.004	0.2934
Hazard-rate (pd+cue)	222.45	7.461	0.003	0.2884

Table 2. Point estimates, percent coefficient of variation (*CV*), and 95% *BCa* smoothed bootstrap confidence limits of perpendicular distances probability density evaluated at 0, $\hat{f}(0)$, expected group size, $\hat{E}(s)$, and abundance of eastern spinner dolphins in the *core* and *North coastal* area of the eastern tropical Pacific ocean. Estimates were computed with data from 1998, 1999, and 2000, using conventional analysis with least squares (*LS*) regression, and with robust (*MM*) regression, and using parametric and nonparametric multivariate (MV) detection functions. $\bar{s}$ is the mean group size, *n* is sample size, *L* total line length in km, and the size of the area surveyed was 6404306 km^2. Parametric MV uses model averaging of $\hat{f}(0)$ (Tables 1A-C).

Year	method	p.d.f. at 0		group size			Abundance	
		$\hat{f}(0)$	% CV	$\bar{s}$ $\bar{s}$ (%CV)	$\hat{E}(s)$	%CV (95% C.I.)	$\hat{N}$ $\hat{N}$	%CV (95% C.I.)
1998 n = 99 L=23911	Conventional + LS	0.3576	8.33	130.6 (12.66)	111.0	15.7 (82.3-149.0)	457277	22.92 (271924 - 682562)
	Conventional + MM				121.5	16.8 (82.5-157.0)	500297	24.42 (286524 - 745214)
	MV parametric	0.3583	8.33		122.3	12.9 (91.1-148.1)	504720	21.75 (292025 - 719462)
	MV nonparametric	0.4271	16.28		103.0	19.0 (72.5-151.7)	506663	24.57 (320963 - 860768)
1999 n = 70 L=17756	Conventional + LS	0.2451	8.00	110.6 (17.22)	83.0	18.95 (60.5- 120.9)	249521	26.25 (144900 - 420512)
	Conventional + MM				107.5	24.63 (71.9- 168.9)	323285	32.52 (177452 - 593821)
	MV parametric	0.2816	11.54		99.1	19.02 (65.4-133.9)	342126	26.91 (204797 - 592131)
	MV nonparametric	0.2911	20.69		82.1	25.59 (46.5-133.4)	293258	35.61 (148824 - 597827)
2000 n = 70 L=17991	Conventional + LS	0.3224	13.79	125.5 (18.74)	108.7	16.95 (79.7-149.5)	422756	26.04 (266149-754605)
	Conventional + MM				125.4	22.05 (94.3-261.3)	487885	31.41 (324160-1281110)
	MV parametric	0.3244	13.79		123.7	18.23 (94.6-195.1)	484174	26.55 (325944-889593)
	MV nonparametric	0.3699	22.86		105.7	17.04 (74.4-146.2)	472026	26.45 (272692-771130)

Table 3. Point estimates, standard errors (SE), percent relative biases (*PRB*), mean square errors (*MSE*) and 95% confidence interval coverages for density of individuals and expected mean group size estimated with different methods. D_{LS} and $\hat{S}_{LS}$ are estimated with univariate detection function and *LS* group size regression on detection probability; D_{MM} and $\hat{S}_{MM}$ are estimated with univariate detection function and robust (*MM*) group size regression on detection probability; D_{COV} and $\hat{S}_{COV}$ are estimated with parametric multivariate detection functions; and D_{KS} and $\hat{S}_{KS}$ are estimated with nonparametric multivariate detection functions. *NB* corresponds to the simple bootstrap confidence intervals and *SB* corresponds to the coverage with smoothed bootstrap method. Data is simulated with an exponential power series detection function, with $W = 5$ km, true group size of 146 individuals, density of groups of 0.95 group/km^2, and density of individuals of 138.7 ind./km^2.

		scale parameters				scale parameters	
		$\alpha_1 = 0.20$, $\alpha_2 = 0.05$	$\alpha_1 = 0.20$, $\alpha_2 = 0.20$			$\alpha_1 = 0.20$, $\alpha_2 = 0.05$	$\alpha_1 = 0.20$, $\alpha_2 = 0.20$
$\hat{D}_{LS}$		136.9	132.98	$\hat{s}_{LS}$		149.4	150.64
SE		0.98	1.03	SE		0.85	0.87
PRB		-1.29 %	-4.12 %	*PRB*		2.33 %	3.18 %
MSE		490.5	568.0	*MSE*		369.6	400.9
coverage	*NB*	88.8 %	92.8 %	*coverage*	*NB*	97.0 %	92.0 %
	SB	90.4 %	94.8 %		*SB*	97.2 %	92.8 %
$\hat{D}_{MM}$		139.8	134.43	$\hat{s}_{MM}$		152.58	152.30
SE		0.99	1.05	SE		0.85	0.90
PRB		0.78 %	-3.07 %	*PRB*		4.51 %	4.32 %
MSE		495.5	575.78	*MSE*		408.2	445.3
coverage	*NB*	94.4 %	94.8 %	*coverage*	*NB*	95.2 %	93.2 %
	SB	95.6 %	93.6 %		*SB*	94.8 %	95.0 %
$\hat{D}_{COV}$		145.5	139.94	$\hat{s}_{COV}$		156.75	156.47
SE		0.89	0.92	SE		0.69	0.69
PRB		4.89 %	0.89 %	*PRB*		7.36 %	7.17 %
MSE		440.55	426.79	*MSE*		350.7	350.5
coverage	*NB*	93.6 %	94.8 %	*coverage*	*NB*	93.2 %	94.8 %
	SB	96.4 %	95.8 %		*SB*	92.8 %	94.4 %
$\hat{D}_{KS}$		143.9	138.56	$\hat{s}_{KS}$		148.9	149.75
SE		1.13	1.17	SE		0.88	0.93
PRB		3.73 %	-0.10 %	*PRB*		2.01 %	2.57 %
MSE		675.2	689.93	*MSE*		394.0	442.4
coverage	*NB*	93.2 %	93.2 %	*coverage*	*NB*	95.2 %	96.0 %
	SB	94.0 %	94.6 %		*SB*	95.2 %	94.4 %

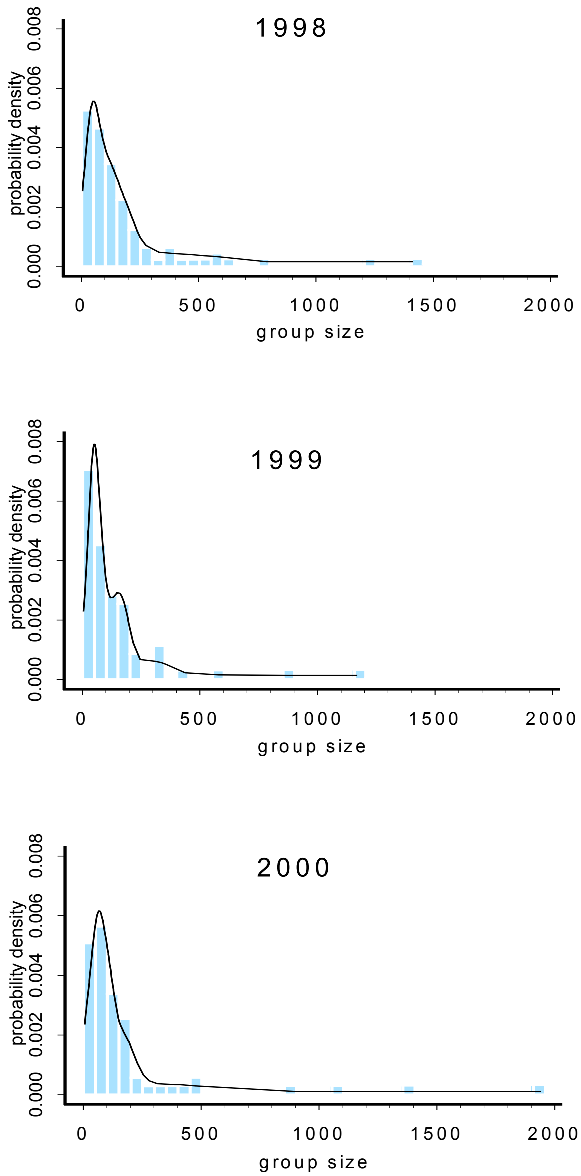

Figure 1. Group size distribution of eastern spinner dolphins (*Stenella longirostris orientalis*) in the eastern Tropical Pacific Ocean during 1998, 1999 and 2000. The fitted line is an adaptive bandwidth kernel smoother density estimate.

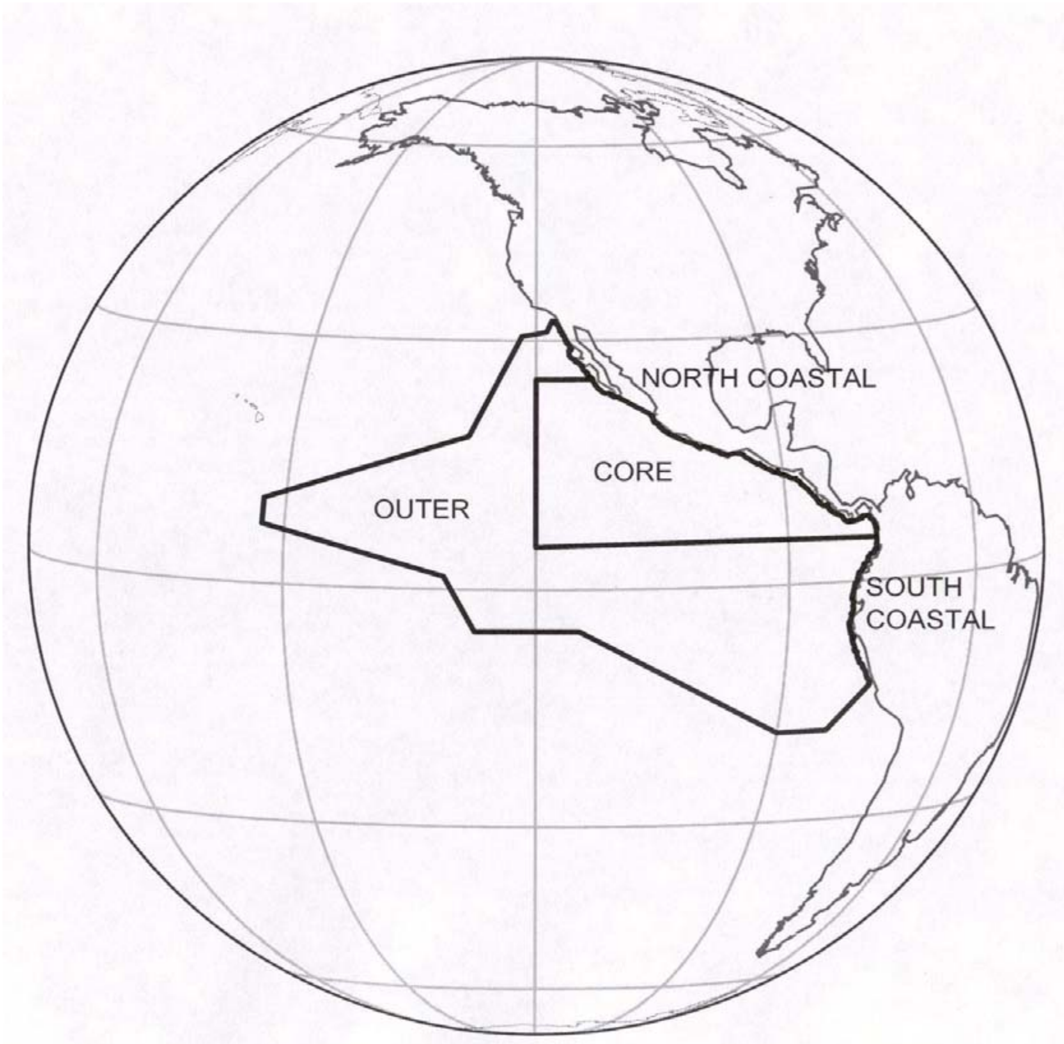

Figure 2. Map of the study areas of the example data set. The total survey area contains four strata, Outer, Core, North Coastal, and South Coastal. For illustration of the methods, the example subset uses data from the Core and North Coastal strata combined.

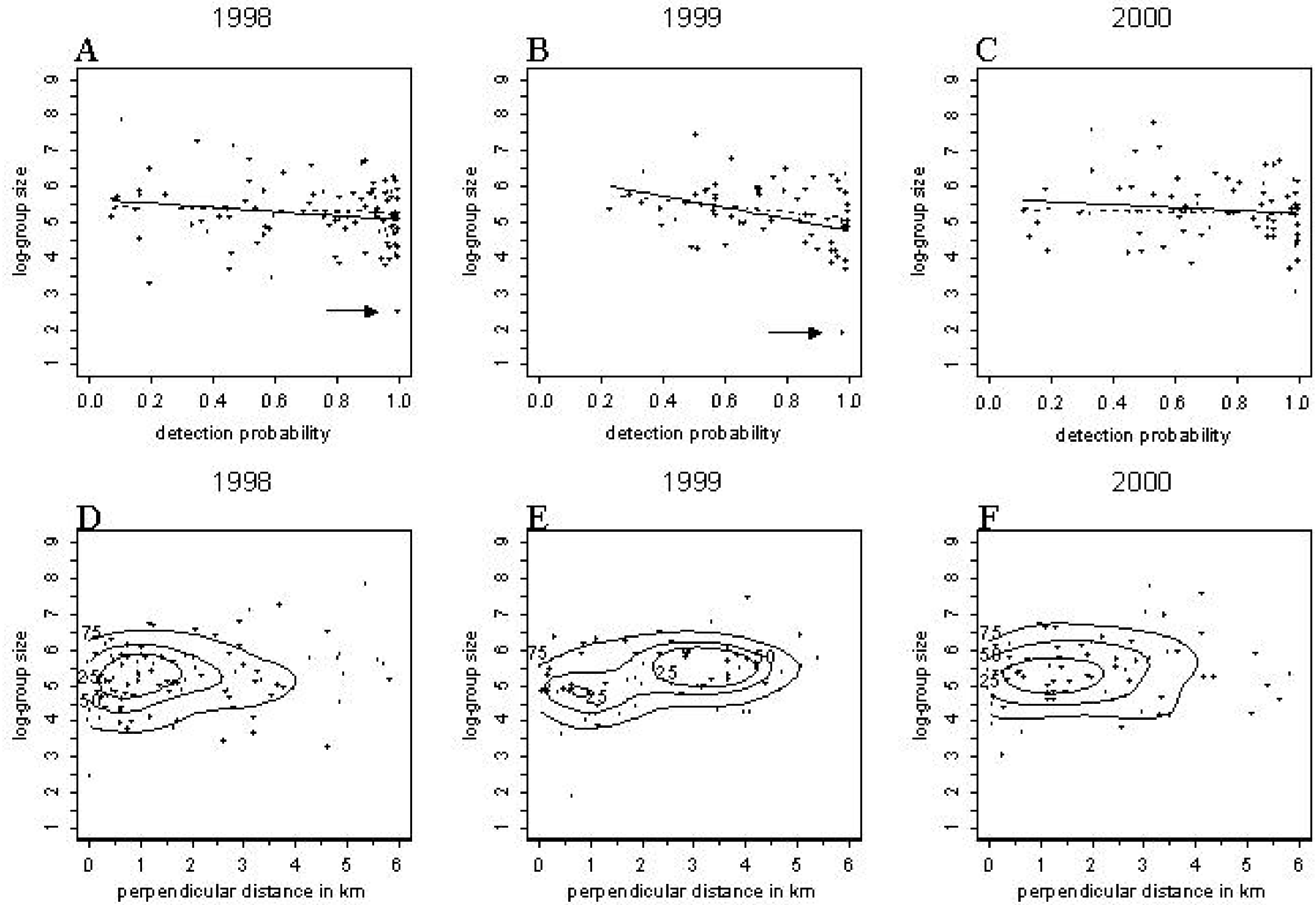

Figure 3. Association between perpendicular distance and eastern spinner dolphin group size on each year as measured by linear regression (A-C), and by a bivariate probability density surface (D-E). Panels A to B show a least squares (continuous) and an *MM* robust regression fit (dashed), with arrows indicating significant outliers. Panels D to E, show bivariate density contours, indicating the percentage of the data points contained.

APPENDIX A

Response to CIE reviewer comments

The manuscript reporting the new analysis tools especially developed for the IDCPA project has been improved according to modifications suggested by reviewers and the final version is included as a Technical report.

Reviewer #1: R. Mohn. Dept. Fisheries & Oceans, Bedford Inst. Oceanography, Dartmouth, N. S., Canada.
Reviewer #2: P. Medley. Consultant, Alne, UK.

Enhancement of Simulations

All the simulations included in the technical report were repeated and additional ones carried out to address the points suggested by the reviewers. The technical report now contains further description of the simulation results. In particular, it is clarified that contrary to R. Mohn's comment, the selection of covariates was to mimic the real data. In the previous simulation tests, main covariates were: perpendicular distance, group size and a two level factor (e.g. bird association). These covariates were commonly selected in the data analysis. Now this can be seen more clearly in the technical report.

Software reliability.-- Given the intensive nature of the computations, a limited number of the simulations were also repeated using different algorithms for maximum likelihood estimates of parametric models. This was the part of the analysis most sensitive to errors. In particular, Buckland's Newton-Raphson algorithm for fitting density functions with expansion series (Buckland 1992) was implemented, and results were compared to the quasi-Newton algorithm in the S-Plus package (MathSoft Inc. 1999), used for the analysis. Results showed similar detection function parameter estimates, and a sensitivity analysis indicated that errors in f(0,c) estimates (probability density function of perpendicular distances and additional covariates evaluated at 0) attributable to the minimization algorithm were minimum, and were similar using both algorithms.

The new analysis, as reported in the technical report, includes parametric model selection and averaging based on the low sample AIC version (QAIC). In practice, this increases computation time a lot but a sensitivity analysis indicates that this method greatly improves the final abundance estimates. Model selection uncertainty in f(0,c) now accounts for the effects of multiple covariate models, and the best combination of models is used to produce averaged f(0,c) estimates. Model averaging takes into account estimates based on conventional univariate models, extensively tested (i.e. program DISTANCE). Our software, especially developed for this analysis, produces the same estimates for these simple models, but also produces estimates using the new covariate models, including multi-model selection and averaging.

Computing the final stock abundance estimates, many minor errors in the code related to different parts of the complex estimation algorithm were detected and corrected. Errors mostly affected the f(0,c) estimates when sample size for estimation was low. In particular, to obtain reliable estimates for the target species, proration of abundance of unidentified species required independent estimates of more than 7 dolphin stocks. The best solution for reducing errors by stock and year was investigated, and the code was implemented and tested accordingly. Tests to detect further errors were run at all moments to ensure the most possible reliable results. In particular, tests on the bootstrap analysis of variability were extensively repeated until consistency in the results between stocks and years was found.

Impact of "rare large events".-- Rare large events are outlier school, of very large size, encountered at large perpendicular distances (pd). The need to impose a truncation pd to fit multivariate parametric models may leave some very large schools out of the analysis, simply because these are detected at the right edge of the truncation distance (W). These may lead to underestimate mean school size. As noted in the technical manuscript the Horvitz-Thompson type abundance estimator requires further truncation than univariate conventional estimators to avoid bias. To decide what is the appropriate truncation distance, the inclusion probabilities, 1/Wf(0,c), should be above 0.1-0.2. In our analysis, all large school sizes could be included, conditional on their detection pd. Further, a sensitivity analysis using the proposed smoothed bootstrap algorithm with school size modelling suggests that the variance was not underestimated, even when very large schools were discarded. Appropriate discussion is included in the technical report.

Correlation among covariates and model

Modelling covariate effects on a log-linear scale did not reduce possible collinearities; however, inclusion of collinear covariates was avoided, based on an *ad hoc* analysis of covariate effects. This analysis was independent, and based on methods similar to those reported in Barlow et al. (2001), with particular examination of matrices of covariances between covariates. Moreover, model selection based on QAIC provides an objective covariate selection. At this point of the analysis, the problem of relationships between covariates was not found to be a source of structural error in models and/or bias, and any potential bias should be small. Future research on the subject should definitely improve the selection process.

Model selection

Model selection and inference was based on multiple models including different sets of covariates affecting delectability by stock and year; however, all the models were based on the half-normal key to provide more consistent estimates for the posterior population modelling. Multi-model selection and averaging was superior in producing consistent results than simply imposing the same models (i.e. same key and covariates across years), because the resulting f(0,c) estimates were sensitive to covariate effects particular to stock and year. Point and variance estimates were improved in that they accounted for model selection uncertainty and reduction in heterogeneity, and model selection based on QAIC reduced the overfit and the chances of structural errors because of low sample size.

Modeling of detection distances.—Medley proposes an alternative method for modelling detection distances, which could be used as a base for future improvements. The method has potential merits, although has great similarities with methods for radial (detection) distance methods proposed in the past. The reliability of these methods has been proven to be much inferior to methods purely based on perpendicular distances (Buckland et al. 2001), and revisit such methods at this point of the present exercise was deemed impractical and unnecessary.

References

Barlow, J., Gerrodette, T. and Forcada, J. 2001. Factors affecting perpendicular sighting distances on shipboard line-transect surveys for cetaceans. J. Cvet. Res. Manage. 3:201-212.

Buckland, S. T. 1992. Algorithm AS 270: Maximum likelihood fiting of hermite and simple polynomial densities. Applied Statistics **41**: 241-266

Buckland, S. T., D. R. Anderson, K. P. Burnham, J.L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling. Estimating abundance of biological populations. Oxford University press, Oxford.

MathSoft Inc. 1999. S-PLUS 2000 Guide to Statistics, Volume 1, Data Analysis. Products Division, MathSoft, Seattle, WA.